

Detailed Comparison of America's Army Game and Unit of Action Experiments^{*}

CASOS Technical Report

Mike Schneider, Kathleen M. Carley, Il-Chul Moon

August 2005

CMU-ISRI-05-139

Carnegie Mellon University

School of Computer Science

ISRI - Institute for Software Research International

CASOS - Center for Computational Analysis of Social and Organizational Systems

Abstract

America's Army game and Unit of Action experiments are simulations of military units. America's Army is a simulation of a squad level military organization, and the Unit of Action experiments simulate the brigade/battalion level command and control structure. Studies of both these domains reveal that configuring and organizing a unit is important in enhancing its performance. Thus, we may argue that there is a fundamental similarity between the game and the experiment. In this paper, we compare the two simulation analyses in terms of their initial features, results, and recommendations. The comparison of the initial features suggests that both domains use similar explanatory variables, and the comparison of the results shows that the findings of these domains also correspond in many ways. Finally, we investigate different ways of capturing implications in both domains, and based on that, discuss future research issues for the Unit of Action experiments such as developing an ideal communication network or policy by adopting a research method used in the America's Army analysis.

^{*} This work was supported in part by DARPA and the Office of Naval Research for research on massively parallel on-line games. Additional support was provided by CASOS - the center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of Darpa, the Office of Naval Research or the U.S. government. Many thanks to John Graham and Craig Schreiber for providing valuable comments

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE AUG 2005		2. REPORT TYPE		3. DATES COVERED 00-08-2005 to 00-08-2005	
4. TITLE AND SUBTITLE Detailed Comparison of America's Army Game and Unit of Action Experiments			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Carnegie Mellon University,School of Computer Science,Pittsburgh,PA,15213			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 23	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Keywords: Organizational Theory, Computer Game, Simulation, Military Experiments, Unit of Action, America's Army, Command and Control

Table of contents

I.	Index of Tables	iv
II.	Index of Figures	iv
1	Introduction.....	5
2	Data Collection	5
2.1	Why we compare America's Army and UA experiments?	5
2.2	America's Army Game Research	6
2.2.1	What is America's Army Game?.....	6
2.2.2	Previous research on America's Army Game.....	6
2.2.3	Collected dataset from America's Army game.....	8
2.3	Unit of Action Experiment Research.....	9
2.3.1	What is Unit of Action Experiment?	9
2.3.2	Previous research on Unit of Action Experiment	10
2.3.3	Collected dataset from UA experiments	11
2.4	Brief comparison of America's Army and UA experiments	12
2.5	Brief comparison of similar measures	13
2.6	Overview of differences in America's Army and UA experiments	14
3	Key measures	14
3.1	Comparison of explanatory variables	14
3.2	Comparison of dependent variables.....	15
4	Comparison between the America's Army and the UA experiment	16
4.1	Comparison of results	16
4.2	Comparison of recommendations	19
5	Recommendation for further analyses	20
6	Conclusion and Discussion	21
	References.....	22

I. Index of Tables

Table 1 The brief summary of basic features of the America's Army game and the Unit of Action experiments	12
Table 2 The summary of comparisons between America's Army team measures and the UA experiment SSA calculation inputs	13
Table 3 Comparison of explanatory variables in the two analyses.....	15
Table 4 Comparison of dependent variables (performance variables) in the two analyses.....	16
Table 5 Two regression analyses from the UA experiment done at Ft. Leavenworth in 2003 and America's Army second data set.....	17
Table 6 Five similar analysis results of America's Army and the Unit of Action experiments....	18

II. Index of Figures

Figure 1 America's Army Game screen shot	6
Figure 2 How-to construct who-talked-after-whom network with the sequence of communication messages	9
Figure 3 Unit of Action experiments	10
Figure 4 Workload Questionnaire Administered During the Fort Leavenworth Exercise	12
Figure 5 A social network and a graph showing dynamic change of the UA experiment at Ft. Leavenworth in 2003	19
Figure 6 A social network from a team in the first America's Army data set and a recommended squad structure based on statistical/social network analyses	20

1 Introduction

Currently, simulation techniques are extensively used in military communities [1]. Simulation has several advantages that cannot be gained from real world experiments. Simulation is cheaper than real world experiments: it is evident that simulation does not require real damage or movement of units or assets. Also, simulation provides a better chance to record information that can be useful for further analyses. Field exercises would be hard to record in a well-organized data format automatically. Though simulation has its own benefits, it should be verified or compared to other experiments or models to make it more concrete [2]. Verification and comparison will capture how well a simulation describes a situation, and also the similarity between two simulations.

Developed by the U.S. Army, America's Army is a game simulating how soldiers fight in battlefields as a squad. In the previous tech reports [3][4], we performed data-mining on the America's Army game log records, and the results from the data-mining gave us some insights that are applicable to the real world. Unit of Action experiment (UA) is a live simulation of a brigade/battalion level command and control (C2) post. The UA experiments were extensively analyzed, and the analyses presented significant insights into enhancing the efficiency of the UA. In the America's Army analysis, we find that explanatory variables and results are similar in both America's Army and the UA experiments. Therefore, in this paper, we will compare America's Army and UA experiments to verify whether the research methodology or suggested guidelines of the former can be applied to the latter.

From the previous America's Army analysis, we can see distinct team organization structures among the top 1000 teams. Furthermore, we can predict the amount of inflicted/received damage based on the network level measures of the team communication/movement network. This demonstrates that the team performance is partly influenced by the team structure, so we proposed an optimal team structure that may be successful in the America's Army game. From the UA experiment analyses, key players, key tasks and key knowledge are identified, and they play important roles in the performance of a UA team. Additionally, we can see that social network distance and physical distance affects the shared situation awareness of a UA command and control post. According to these findings, we can see that organizational design and social network analysis are commonly important in understanding the two research venues. Thus, we will compare the two research domains to discover new data collection strategies, research methods and implications that are applicable to each other.

2 Data Collection

2.1 Why we compare America's Army and UA experiments?

In this report, we compare America's Army analysis and UA experiment analyses. The two domains are different from each other in some aspects such as the layer of simulating military units, the size of the simulation participants and the ways the two simulations proceed. It would be insightful to know how researchers analyze the two different simulations. If we see the common analysis method in the two analyses, the analysis method may be applicable to many other simulations. Also, it has been known that the two analyses concentrate on the social

networks and their influence to teams' performance. Therefore, we try to answer how social networks of both domains are analyzed and what types of recommendations are made from them. If there is a type of recommendation that is suggested in one domain, it would be possible to make a similar type of recommendation for the other domain by using an equivalent analysis method.

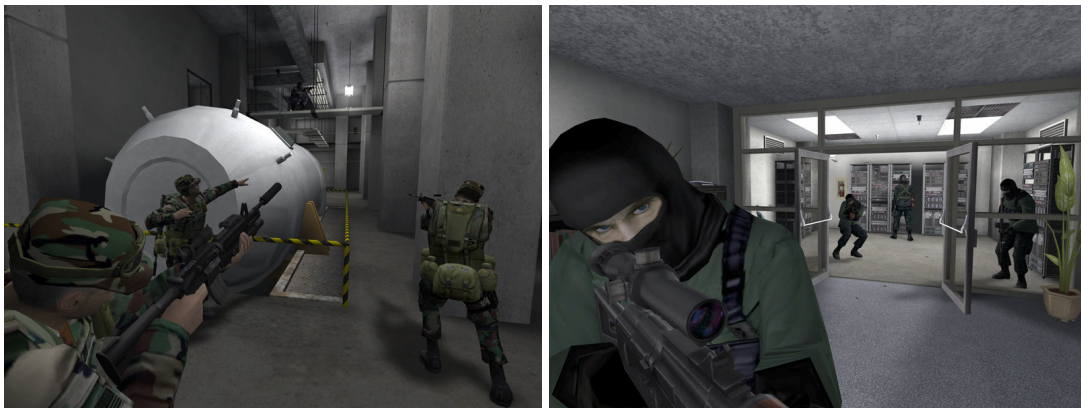
2.2 America's Army Game Research

2.2.1 What is America's Army Game?

The online multi-player computer game America's Army, has more than three million registered players. Developed by the U.S. Army, the game was designed as a recruiting and training tool to paint a realistic portrait of combat in the U.S. Army. The game falls into a first person shooting (FPS) game genre, and all the game features are based on the real world. The game is the duel of two teams, usually an assault team and a defense team, and a team consists of one to fourteen players. The team can win the game by killing all of the opposing players, or accomplishing the goal for that mission, such as securing an oil pipeline, crossing a bridge, etc.

Though the original role of America's Army is about the recruitment of young adults, it is also possible that we can learn lessons from the game play because its features are based on the real world. Furthermore, it can be used as a training tool for combat soldiers to give them virtual experience of future battle fields. When we consider these potentials of America's Army, it is meaningful to compare America's Army and the real world to obtain lessons from its log records and to improve it to be used as a useful training tool.

Figure 1 America's Army Game screen shot



2.2.2 Previous research on America's Army Game

After the release of America's Army, there were a number of research papers published about the game. These papers can be divided into two categories: evaluating America's Army as a tool for stimulating recruitment, and as a tool for training inexperienced soldiers. The research done by Belanich et al [5] is typical of research on America's Army's usefulness as a recruiting tool.

Belanich surveys 21 experiment participants about the information presented during the game and motivational aspects of the game before and after playing America's Army. The assessment of motivational features suggests that PC-based training games should be designed with attention to challenge, realism, control, and opportunities for exploration, and America's Army should be improved in those areas. Also, the paper written by Nieborg [6] explores four aspects of America's Army: Advergame, the integration of advertising messages in online games; Propagame, a strategic communication tool; Edugame, a tool for introducing people to the goals and values of the Army; and Test bed & tool dimension, an experimental test bed and tool.

Other researchers have looked at America's Army as a part of a framework for training soldiers. Farrell et al [7] shows how America's Army has been used at West Point to practice land navigation in a virtual environment, but the usage of America's Army was limited to the test environment for the evaluation of cadets, not actual course work. Little research has been conducted concerning America's Army as a training tool which incorporates analysis of the game play elements, players' behavior, and so on. Also, with our best knowledge, there was no significant research about comparing America's Army and UA experiments.

The first America's Army tech report [3] researched the log record dataset at player level, team level, and clan level. Particularly, many statistical methods are applied to discover traits of social networks, based on Report-In communication networks, of the winning teams in America's Army. From the research, several commonalities among the top teams were found, and some outlying teams were adopting unusual ways to win. Among these traits of the top teams, we could see that the received/inflicted damage amount can be predicted by using the ORA measures [8] of communication social networks and the high clanishness representing the common background among team members is the one of the traits of winning teams. Because the UA experiments reveal the importance of communication distance [9] among participants and the common background experience and education, we could conjecture there would be some similarities between two domains because findings from both analyses are similar.

The second America's Army tech report [4] could reveal another similarity that couldn't be found in the first tech report. Because we could get the event location coordinates in the second tech report, we found out the how the virtual proximity between team members affect the team performance. This proximity measure is similar to the physical proximity of the UA experiments [10]. Analyses suggested that a team performance would be improved if the team members stay closely, and the suggestion corresponds to the UA experiments analysis results that present lower physical proximity (closer physical allocation among players) will enhance the team performance. Not only these additional findings, but also we did some qualitative comparisons between two domains in terms of explanatory variables, performance measures, level of units simulated by the experiments, etc.

The America's Army journal paper [11] analyzed the features of America's Army to investigate its potential value as a training tool for inexperienced soldiers. We looked at the realism of the game itself, in terms of how well it corresponds with the real world, and we looked at the behavior of high-performing players within the game, to see if the strategies they adopted corresponded to the behavior of real soldiers in combat. We analyzed the first and the second log record dataset at the same time, and we surveyed previous research about squad-level infantry

units to determine how well the two correspond. The realism of America's Army is verified from three viewpoints: weapons, communications, and rules of engagement. The similarities between winning players and trained soldiers are investigated in terms of: weapon usage, communication usage, and team structure. Comparisons between America's Army and the real world revealed a number of similarities and the actions of winning soldiers and trained soldiers are almost identical. Finally, we identify some improvements that would further increase the America's Army game's usefulness as a training tool.

2.2.3 Collected dataset from America's Army game

Two datasets were collected from America's Army game servers. The first dataset is collected over 200 servers over the course of 14 days. The second log record data was recorded off of 138 America's Army game servers over the course of 23 days. We primarily use the second dataset in this paper because it provides the physical location information of every event unlike the first dataset. Each line of the log files represents one event recorded by the servers. Every log records in the second dataset have three coordinates information representing a point on the 3D. These events describe the game statistics, where "game" is the unit for the data analysis. Each *game* contains two types of events: *logging events* and *collection events*. The logging events describe the teams and the players, the collection events represent actions performed by players. There are seven types of events used for the data analysis:

1. *Team is initialized*
2. *Player enters the team*
3. *Weapon is used*
4. *Damage caused by the weapon*
5. *Communication between the players*
6. *Player leaves the team, scores are reported*
7. *Team finishes, outcome is recorded*

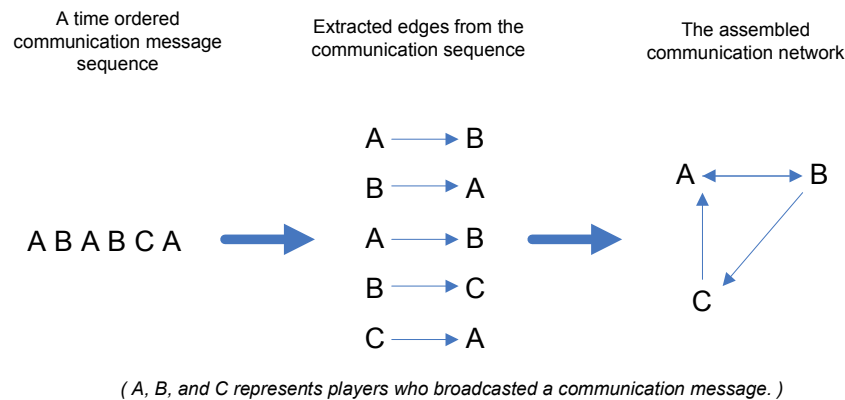
There are always two teams per game playing against each other. A team can have up to 14 players. The logging event *team finishes, outcome is recorded* contains information of either the team wins or loses the game, as well as the initial and final number of players. The logging event *Player leaves the team, scores are reported* has multiple measures of the performance in the game, individual scores: leader score, wins score, objectives score, death score, kills score, ROE score, and total score. Aggregate scores can be calculated for the whole team if one aggregates the scores of the individual players playing in the team. Similarly, weapon usage and damage can be aggregated for the whole team. Some portion of the data files ended abruptly without logical ending for the games, which caused some games to miss events of one or more types mentioned above. Also, some parts of the dataset caused another parsing problem sporadically. There were cases that the location information is not recorded at the end of event log records or location information is recorded all zeros. These cases were discarded, too. However, most of the dataset was well formed and concrete because it was automatically recorded by game servers.

The social network of a game is created by two heuristics. When it comes to the communication social network, the 'who-talked-after-whom' heuristic was used (refer to figure 2). When we

created the physical social network, we tested two conditions, temporal constraints and physical location constraints, with every pair of events and identified edges of the physical social network. Below is the detailed description of the two conditions.

- **Timing Condition**
 - 1) Calculate the average game length of whole dataset
 - 2) Assume that the one-tenth of the average game length would be good time frame to determine whether two players' event happened at the same time or not.
 - 3) Player A's event time \leq Player B's event time
 Player B's event time $<$ Player A's event time + 0.1 X (Average game length)
 \rightarrow Player A and Player B acted at the same time.
- **Location Condition**
 - 1) Calculate the average scatterness of whole dataset
 - 2) Assume that the average scatterness would be a good standard distance to cooperate together.
 - 3) Distance(Player A's event, Player B's event) $<$ (average scatterness)
 \rightarrow Player A and Player B acted at the same place.

Figure 2 How-to construct who-talked-after-whom network with the sequence of communication messages



2.3 Unit of Action Experiment Research

2.3.1 What is Unit of Action Experiment?

The Unit of Action (UA) is the unit that will replace the current force battalion in the objective force. The objective force [12] means the future army that the current U.S. Army wants to be in the future. The UA is the key army organization that is deployable in short time and transportable with tactical transport airplanes, so its characteristic will be strategic and operational responsiveness, rapid deployment, and ability to change patterns of operations. To meet those requirements, the Army has developed the Future Combat Systems (FCS), the Unmanned Ground/Air Vehicles (UGV/UAV), the Stryker Infantry Vehicles, the Land Warrior

System (LW), etc. Not only the above hardware improvements, but also software and organization improvements has been investigated and implemented.

For the organizational improvements, the Command and Control (C2) system of the UA is investigated with three experiments, Experiment #1: Ft Leavenworth Kansas, October-2003; Experiment #2: Ft Lee Virginia, February-2004; Experiment #3: Ft Knox Kentucky, June-2004, and they are the UA experiments. For example, the UA experiment [10] at Ft. Leavenworth KS in June 2003 is done with fifty-six army officers for four days. The army officers served as role players for an experimental C2 and performed their experimental duty with a computer scenario. The role players were brought together specifically for the purpose of conducting this exercise and they did not know each other before the experiment. The experiment participants were assigned to a functional cell that consisted of three to eight participants. Each cell gathered and analyzed necessary information and entered battlefield actions based on the information and the analyses. Because of physical allocation, cell members could talk to the other members, but they could not directly communicate with the participants assigned to the other cell. During the experiment period, each participant submitted questionnaires asking communication contacts, tasks, and information about the previous work, and the questionnaires are used to analyze and assess how the experiment went.

Figure 3 Unit of Action experiments



2.3.2 Previous research on Unit of Action Experiment

The main stream of the previous research [9, 10, 13] on the UA experiment is the analyses of its command and control (C2) structure efficiency. Because the UA experiment has many officers cooperating and communicating to each other, well-established C2 has been considered as a key to the success of the UA experiment. To evaluate the C2 structure, the UA experiments analyses utilized Shared Situation Awareness (SSA), and it has become one of the most convincing metrics to measure the C2 efficiency.

Endsley [14] defines that SSA is the degree to which team members possess the same SA on shared SA requirements. With the definition, the paper argues three important elements of SSA that are Shared Information (SI), SA devices and SA mechanisms. Those elements are carefully treated during the UA experiments, so the communication methods and shared information that

are relevant to the three elements are collected during the experiments. Additionally, this paper introduces Team SA processes and the differences in the processes of effective teams and ineffective ones. The introduction of the Team SA processes influenced the UA experiments settings and their analyses.

After Endsley's introduction of SSA, there were discussions about it in military communities. Among many papers discussing it, Nofi [15] tried to build a way to measure SSA in a team. In the paper, he said that it is possible to compose a systematic definition and to develop objective approaches to studying the process by which SSA arises. However, he failed to show any specific formula or evaluation methodology for SSA. On the other hand, Graham [16] experimented the efficiency of the UA C2 structure for three times, and the series of experiments are the UA experiments that we are discussing in this paper. Because the experiments used SSA as a performance measure, he could discover important elements of SSA through the analyses of the experiments. First, Graham et al [9] discovered there are evident links between SSA and social network distance, physical distance. In the paper, he did a regression analysis with physical distance and social network distance against the mental model congruence that is related to the SSA. He could see that the two variables can be predictors for it. After the research, he wrote a dissertation [16] based on the three UA experiments and measuring SSA of the experiments. In the dissertation, he derived three explanatory variables, physical distance, social network distance and background similarity, from prior research papers. Then, he proposes a statistical way to calculate the SSA with these three variables. The UA experiment information we are using in this paper largely comes from his dissertation, and the SSA measure in this paper follows his definition and calculation formula.

2.3.3 Collected dataset from UA experiments

Three UA experiments are done for the investigation of UA's command and control structure. The three experiments are Experiment #1: Ft Leavenworth Kansas, October- 2003; Experiment #2: Ft Lee Virginia, February-2004; Experiment #3: Ft Knox Kentucky, June-2004. Due to the circumstances of the experiments, there are no experimental and control groups. Only one or two UA command posts are simulated in each experiment by experienced army officers. After the simulation is started, it is periodically stopped to collect observations from the participants of the experiment. Technically, the data collection was done by a computer program. The automated self-report collection system was executed every 60-90 minutes during the simulation, and the simulation was frozen until all the responses were collected. The self-report somewhat resembles to the NASA TLX form (refer to figure 4), but we collected more data than the NASA TLX form specifies. For example, we could construct a social network between participants based on the frequency of coordination that is collected from the participants' responses.

Figure 4 Workload Questionnaire Administered During the Fort Leavenworth Exercise

Form1

Workload Questionnaire

Low High

1 2 3 4 5 6 7

Please answer the following questions about WORKLOAD.
Rate Each Question From 1 to 7 (1- lowest, and 7 - highest).

RATE YOURSELF

- 1 **MENTAL DEMAND** How much mental and perceptual activity was required during the last hour(e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
- 2 **TEMPORAL DEMAND** How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
- 3 **EFFORT** How hard did you have to work (mentally and physically) to accomplish your level of performance in the last hour?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
- 4 **OWN PERFORMANCE** How successful do you think you were in accomplishing the goals of the task set by yourself and your role?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
- 5 **FRUSTRATION LEVEL** How discouraged, irritated, stressed and annoyed did you feel during this round?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
- 6 **PHYSICAL DEMAND** How much physical activity was required (e.g., mouse, keyboard, controlling, activating, etc.)?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Page 1 of 14

2.4 Brief comparison of America’s Army and UA experiments

In the previous tech reports, the UA experiments [3, 4] are briefly compared to the America’s Army game. (Refer to Table 1) First, the sizes and levels of the simulated military units are different. America’s Army is a simulation for squad unit consisted of one to fourteen members. UA experiments simulate brigade/battalion level command posts consisted of over fifty army officers. Because the types of units are different, the information and the action that participants face are different. The events and the actions in America’s Army are dynamic and changing, and those in the UA experiments are stable and well formed.

However, there are still similarities between two experiments. Because both simulations deal with the social network analysis, ORA measures [8] could be applied to evaluate and extract the attribute of the social network for each experiment trial. Also, measurement for unity of unit and congruence of the knowledge might be good measures for the social network. Not only the similarities in explanatory variables, but also the effort to predict the performance measures of both domains would be a good similarity. Furthermore, the possibility that we can predict performance measures in both domains with similar explanatory variables emphasizes there are similar aspects in being successful in both experiments and it also might be possible to apply insights obtained from one domain to the other domain.

Table 1 The brief summary of basic features of the America's Army game and the Unit of Action experiments

	America’s Army	UA experiments
Size/Level of Unit	Squad level soldiers	Brigade level staff officers
Performance measures	Received/Inflicted	Shared Situation Awareness

	Damage, winning/losing	
Explanatory variables for the performance measures	ORA measures of social network among soldiers / Clannishness	Physical distance, social distance, and background similarities
Can performance measure be predicted by explanatory variables?	Yes	Yes

2.5 Brief comparison of similar measures

The performance measure of the UA experiments is SSA, and SSA is calculated from three inputs: agent-to-agent interaction/communication; agent-to-agent physical proximity; and agent-to-agent socio-demographic similarity. In America's Army, three measures that are similar to above three inputs were also calculated.

$$SSA(i, j) = \alpha e(i)e(j) + \beta P(i, j) + \frac{\delta S(i, j)}{\gamma G(i, j)} + \mu A(i, j)A(j, i)$$

Let $\alpha, \beta, \gamma, \delta, \mu$ be Real number

Let A = Agent X Agent interaction/communication matrix

Let P = Agent X Agent physical proximity matrix

Let S = Agent X Agent social demographic similarity matrix

Let e = eigenvector centrality measure computed on A

Let G = dyadic geodesics computed on A

Above SSA calculation formula shows that the increment of the three variables, interaction/communication, physical proximity, social demographic similarity, will increase the SSA. Also, it was commonly believed that the higher SSA will result the positive outcomes of the UA experiment. In America's Army, the higher frequency of Report-In is one of the most noticeable characteristics of the winning teams. Also, the lower scatterness (higher physical proximity in virtual space) is one of the other characteristics of the winning teams. Finally, the higher clanishness can be seen in the winning teams. Therefore, when we consider the above similarities, we can speculate that the winning team in America's Army might have higher SSA than the losing teams have. Thus, with the similar requirements of high performance in both domains, we would be able to conclude that SSA would be good performance measure not only for the UA experiments, but also for America's Army.

Table 2 The summary of comparisons between America's Army team measures and the UA experiment SSA calculation inputs

America's Army statistical result	SSA inputs
-----------------------------------	------------

Number of communication (Number of Report-In)	interaction/communication
Scatterness	physical proximity
Clanishness	socio-demographic similarity

2.6 Overview of differences in America's Army and UA experiments

First of all, it should be noted that the UA experiments are for higher layer military command and control structure such as battalion/brigade command posts, and America's Army is the simulation for the most bottom layer military unit such as special force squads and army squads. Therefore, the numbers and the attributes of interacting agents are very different from each other. For example, while the participants of the UA experiments rely on more stable, well-organized information, the America's Army gamers, on the other hand, face dynamic, changing information for virtual combats between their own squad and an enemy squad. What the UA role-players should decide are plans and reactions to particular situation, and what the America's Army gamers should plan are moving their avatars in a virtual battlefield and shooting their virtual weapons. The intrinsic differences of layers and tasks in both domains make them incomparable in some sense.

Furthermore, both experiments adopt different types of performance measure: winning, scores defined by the Army and received/inflicted damage for America's Army and shared situation awareness (SSA) for the UA experiments. Because America's Army is an on-line game, every movement and activity players do can be recorded and scored by a system. Therefore, determining which team wins, how much a player received shooting, etc are very obvious, and those measures can be easily used as performance measures. However, in the UA experiments, it is quite hard to track all the action role-players make, and it is also difficult to assess the role-players performance. Therefore, the SSA based on the questionnaires completed by the participants would be one of the possible performance measures we can obtain.

Finally, the collected datasets are different. In America's Army, every action and communication will be recorded by a computer program, so we can get well formed datasets for every action players make. Every record for an action includes virtual location of the action, so it is easy to analyze its physical factors. On the contrary, the UA experiments are done in the real world. Therefore, we cannot have an automatically recorded log like the America's Army dataset. All we obtain from them are questionnaires collected after a certain phase or an action is done. Thus, there are intrinsic differences between two datasets.

3 Key measures

3.1 Comparison of explanatory variables

We first compare the used explanatory variables in the two analyses. In table 3, we divided the variables in five categories. Due to the focus on the social network, both analyses include social network measures such as average distance, in/out degree centralizations, span of control, etc. It is also noticeable that both created social networks in two perspectives: communication and physical location. The network measures are calculated by ORA with the two social networks for

each domain. Additionally, measures for background and experience similarities are considered as variables. Clanishness strong means the ratio of same clan members out of team members, and clanishness weak implies that the ratio of clan participating members out of team members. These clanishness measures resemble to the status similarity and organization similarity.

On the contrary, there are variables that are calculated only in America's Army analysis. They are mainly resource related variables. We conjecture that this resource related measures were easy to calculate because America's Army provided a log record that is detailed and well formed. For the UA experiments, it is recommendable to analyze these resource related variables in future analysis if possible. For instance, the numbers of information experienced officer, operation experienced officer, used resources for the operation, etc will give us valuable insights for the UA experiments.

Table 3 Comparison of explanatory variables in the two analyses

Category	America's Army	UA experiments
Communication	Communication network level measures (by ORA)	Communication network level measures (by ORA)
	Number of Report-In	Social network distance
	Number of Commo.	
	Number of Normal Comm.	
Movement (Physical location)	Movement network level measures (by ORA)	Physical location network level measures (by ORA)
	Scatterness	Physical Proximity
Background	Clanishness strong	Status Similarity
	Clanishness weak	Organization Similarity
Experience	Average experience	Command Experience Similarity
Resource	Number of weapon types	
	Number of weapon fire	
	Heavy weapon presence	
	Medic presence	

3.2 Comparison of dependent variables

The performance variables in the two analyses are quite different from each other. America's Army has more than ten performance variables. Winning, total score, leader score, etc are already specified in the given dataset, and received damage amount, inflicted damage amount, survival ratio, and kill ratio are calculated based on the dataset. New score is the overall score that gets input from some of the given scores and some of the calculated measures. The below formula is the formula for new score calculation.

$$\text{New_score} = a_0 + a_1 * \text{score_leader} + a_2 * \text{score_wins} + a_3 * \text{score_goal} + a_4 * \text{score_death} + a_5 * \text{score_kills} \\ + a_6 * \text{score_roe} + a_7 * \text{survive_ratio}(\text{friendly_players}) + a_8 * \text{survive_ratio}(\text{enemy_players})$$

The performance measure in the UA experiments is SSA. Unlike the dependent variables in the America's Army analysis, the SSA is calculated from the explanatory variables. It is questionable that whether a performance measure should be calculated based on log records directly or it should rely on the explanatory variables setup by the analysis. However, because there is no automatic event log record in the UA experiments, it seems that SSA is good alternative way to measure the efficiency of the UA experiments.

Table 4 Comparison of dependent variables (performance variables) in the two analyses

Category	America's Army	UA experiments
Performance measure existing in the datasets	Winning	
	Total score	
	Leader score	
	Wins score	
	Goal score	
	Death score	
	Kills score	
	ROE score	
Performance measures calculated with the datasets	Received Damage	
	Inflicted Damage	
	Survival Ratio	
	Kill Ratio	
Overall performance measures	New score	Shared Situation Awareness

4 Comparison between the America's Army and the UA experiment

4.1 Comparison of results

Considering the similarities between two research domains, we tried to compare the regression analysis result on the performance measures in both domains more extensively. Unlike the UA experiments have one performance measure, Shared Situation Awareness (SSA), the AA analysis has multiple performance measures, so we choose winning as a performance measure by setting winning as one and losing as zero for the comparison because it is the most fundamental performance measure of the AA analysis. Table 5 shows two regression analysis results came from the AA analysis of the second data set and the 2003 UA experiment held at Ft. Leavenworth. According to the R-squares, approximately 20% of the UA performance measure can be predicted, and 7% of the AA performance measure can be explained. More importantly, we could see some tendencies by comparing coefficients between two regressions. For example, organization, status and command experience similarity in the UA experiments is similar to clanishness strong in the AA analysis; social network distance to the frequency of various types of communication; and physical proximity to inversed virtual space proximity. Also the matching variables have same tendencies in terms of the coefficients. Higher back ground similarity, closer social/communication network distance and closer role-players allocation will be preferable to increase the performance measure in both domains. According to these same tendencies in the coefficients of the regression analyses, we could see that some guidelines, such as close and frequent communication activity, close unit location and similar role-player

backgrounds, to be successful in America's Army can be applied to the UA experiments, and vice versa. Furthermore, with these similar tendencies in both domains that are different in hierarchical layers, size of organization, and task/objectives of organization, we might be able to conjecture that those tendencies might be fundamental ways to improve military organization performance. However, we could not compare the standardized coefficients from both domains. Thus, it may not be possible to order the influence power of each coefficient, but still we can compare the sign of a coefficient because the standardized coefficient will have the same sign of the un-standardized coefficient.

Table 5 Two regression analyses from the UA experiment done at Ft. Leavenworth in 2003 and America's Army second data set

	UA Experiment		AA Analysis	
Variable Category	Variable	estimate	Variable	estimate
Intercept	Intercept	0.460	intercept	0.406
Time Flow	Session	-0.005		
Physical location	Physical Proximity	-0.181	Virtual space proximity	0.008
Comm. level	Social Network Distance	-0.070	Num. of ReportIn	0.250
			Num. of Commo	0.116
			Num. of Normal comm	-0.131
Background/Experience Similarity	Status Similarity	0.044	Clanishness strong	0.045
	Organization Similarity	0.031		
	Command Experience Similarity	0.023		
	r-squared	0.200	r-squared	0.297

Not only the similarities between the regression analysis results in America's Army and the UA experiments, some other analyses results correspond to each other. We summarize the similar analyses results in table 6. We could find five similar aspects: operation, cell/team composition, communication style, background/experience similarity, communication trends when a cell/team is immature.

- Operation: Inherently, the Unit of Action experiments emphasize how to coordinate operations among officers, so it is quite obvious that each cell has its own plans and operational components. On the other hand, as a squad level simulation, America's Army demands tactical plans and decisions from teams. When we consider that the teams having high clanishness measure do well in the games, we conjecture that they discuss those tactical plans and decisions in advance as clan members.
- Cell/team composition: Each cell of the UA experiments has three to eight assigned officers and they should operate properly as a specific role-player. Therefore, it is quite natural that officers having knowledge about the role would do better than other officers who have no previous experience. Also, it is confirmed that there are medic specialized top players in America's Army, and they pick their career path from the beginning and keep selecting the medic role.

- **Communication Style:** During the UA experiments, there is a CUB that gives cells a chance to organize their Report-Ins, and it is identified that the after CUB the UA performance improves. The winning teams of America's Army do send out Report-In frequently, and we could see some specific Report-In network shapes adopted by the top teams.
- **Background/experience similarity:** The officers of a UA cell have to spend some time to maximize their performance. This time might give them opportunities to know each other's working style and specialties. The winning teams of America's Army show high clanishness. It means that teams having many team members in a same clan do better than other teams.
- **Communication trends when a cell/team is immature:** At the start of the UA experiments, the experiment participants interact to know each other's role, task, and knowledge. When an America's Army team has low clanishness, the team might spend more time on Normal communication that is similar to verbal communication to coordinate their plans and know each other's play style.

Table 6 Five similar analysis results of America's Army and the Unit of Action experiments

	Unit Of Action experiments	America's Army games
Operation	A cell has a strong planning and operational component.	If a team has a plan for game play, it is likely to be discussed beforehand: predominantly operational – planning is offline (possibly in clans)
Cell/Team composition	Cells where specialized functions (e.g., medical or intel) were played by inexperienced personnel - Had lower shared situation awareness - Had lower performance	Teams where specialized functions (e.g., medical) were played by inexperienced personnel - Had lower estimated situation awareness - Had lower performance
Communication Style	Over time the unit performance and shared situation awareness dropped until a CUB was held at which the commander discussed the situation and what was important and sought information from each unit lead. Then performance improved. There was no organized report in for all units prior to this meeting where they had to listen to each other's reports.	Over time teams that had regular organized report in - Had higher estimated situation awareness - Had higher performance

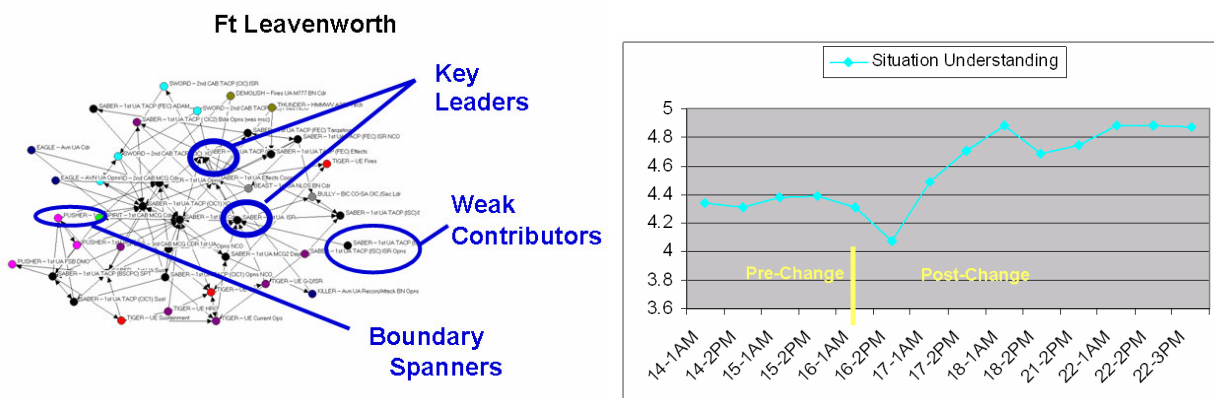
Background/Experience Similarity	Unit has to work together for a while before they exhibit high performance	Teams with more members who have played together before or were in same clan exhibit high performance
Communication trends when a cell/team is immature	Initial interaction is all about trying to understand each other's roles	Teams with inexperienced players spend time in dialogs learning about other's roles – much dialog may also be occurring offline or through clans.

4.2 Comparison of recommendations

There is a common goal between America's Army game analyses and the UA experiment result analyses: obtaining insights of how a military unit is organized and what the enhanced future organization is. To pursue the objective, both analyses have used social network analysis approaches very extensively, and the research methodology induced a similar research process. First, we constructed social networks based on the given data, do statistical analyses of the social networks and their network measures, and make implications based on the analyses results.

Each UA experiment produces one social network having all the experiment participants. By analyzing the social network, we investigate who knows more, who acts more, and who the important actors are. Figure 5 shows a typical social network obtained from the UA experiment done at Ft. Leavenworth in 2003. As the figure represents, we can identify who key leaders/weak contributors/boundary spanners are. Furthermore, we can observe the dynamically changing social network and evaluate how much the unit has situation understanding at certain moments. These are the general analysis procedure from the viewpoint of dynamic social network. By using these analysis results, we can enhance the SSA of a unit by restructuring and reconfiguring the unit.

Figure 5 A social network and a graph showing dynamic change of the UA experiment at Ft. Leavenworth in 2003



America's Army game research deals with not one social network but over hundreds of thousands social networks because each team has its own social network unlike the UA

The optimal structures from America's Army analysis have the credibility to some extent because they are obtained by analyzing the top 1000 teams out of 150,000 teams. On the other hand, we cannot make such an optimal structure for UA experiments if we have small number of experiment logs. Thus, it is recommendable to repeat the UA experiments more and record their communications in detail. With more UA experiments, we will be able to utilize other statistical analysis methods to obtain valuable insights for its success.

Because America's Army is a computer game, it can record every communication messages. However, the UA experiments do not record all the communication messages and do not specify types of messages, such as acknowledgement, Report-In, normal communication, etc. It is revealed that different communication types contribute team performance differently in America's Army. Therefore, it would be useful to record communication message frequencies and types during a UA experiment and analyze them as we did in America's Army.

Also, we should make more performance measures for UA experiments. America's Army analysis uses four performance measures mainly, such as received damage, inflicted damage, new score and winning. We conjecture that UA experiments might setup more performance measures and record them, i.e. response time, overall score and event outcome.

6 Conclusion and Discussion

America's Army game and Unit of Action experiments are simulations of military units though they target different levels of them. America's Army is a simulation for a squad level military organization. Each team of the game consists of one to fourteen players, and the game lets the players experience what will happen in a real battlefield. On the other hand, the UA experiments simulate the brigade/battalion level command and control structure with over fifty army officers. Because America's Army is a virtual game, its software can figure out how much a team scored and which team won. The UA experiments were done in the real world, so we have to rely on questionnaires to evaluate the workload and the SSA.

Furthermore, the America's Army analysis result may not directly applicable to the UA experiments. First, the team size of America's Army is much smaller than that of the UA experiments. Many network level measures used in the social network analysis may be sensitive to its network size. Second, the organization dynamics in America's Army teams and the UA experiments are different. We conjecture that an organization may prefer different organizational structures if it faces different situation and consists of different types of role players. Besides of the two difficulties, numbers of experiment trials, contents of the datasets, difference in the performance measures make it difficult to apply insights from one domain to the other.

However, we can see a fundamental similarity between these simulations because the research of both these domains reveals that configuring and organizing the unit and the team are important in enhancing their performance. Specifically, explanatory variables for performance measures in both domains are equivalent. In both analyses, the social network measures, background/experience similarity and physical proximity are major factors in explaining and analyzing the performance of a team or a unit. Furthermore, the same analysis methodology is applied to both domains. For example, regression analyses are done with similar explanatory

variables and performance variables. The results of the regression analyses are similar for America's Army and UA. We can see that equivalent explanatory variables affect the performance measures of each domain similarly. Frequent communication helps to increase the SSA in the UA experiments and may lead to winning in America's Army. High background/experience similarity enhances the SSA and the winning ratio. So does close physical allocation.

Despite of the similar analysis methods and results, the two investigations suggest different recommendations, and we think future UA experiments will give us more significant insights by proposing an optimal UA C2 structure. So far, the UA experiments concentrated on formulating a systematic way of measuring the SSA, and it was successful. According to Graham's dissertation [16], the SSA evaluation formula is made and tested with three UA experiments. Therefore, we may focus on creating optimal UA C2 structures based on the SSA evaluation formula in future analyses. First, we will evaluate each UA C2 structure based on the SSA. Second, we find ideal network level measures from the evaluation: we can average the network level measures of several top UA C2 structures. Finally, we reconstruct an optimal UA C2 structures based on the averaged network level measures. This procedure is used in America's Army, and it may also be applicable to the UA experiments because both are dealing with social network information of both domains.

References

- [1] Roger D. Smith (1998), Essential Techniques for Military Modeling & Simulation, Proceedings of the 1998 Winter Simulation Conference, Washington, DC
- [2] Robert G. Sargant (1987), An Overview of Verification and Validation of Simulation Models, Proceedings of the 1987 Winter Simulation Conference, pp 604-611, Atlanta, GA
- [3] Carley, K. M. , Moon, I. , Schneider M. , and Shigiltchoff O. (2005), Detailed Analysis of Factors Affecting Team Success and Failure in the America's Army Game, Tech Report, CMU-ISRI-05-120, CASOS, Carnegie Mellon, Pittsburgh PA.
- [4] Moon, I., Carley, K. M. Schneider M. , and Shigiltchoff O. (2005), Detailed Analysis of Team Movement and Communication Affecting Team Performance in the America's Army Game, Working Paper, CASOS, Carnegie Mellon, Pittsburgh PA.
- [5] Belanich, J. , Sibley, D. E. , and Orvis, K. L. (2004), Instructional Characteristics and Motivational Features of a PC-based Game, Research Report 1822, U.S. Army Research Institute for the Behavioral and Social Sciences
- [6] Nieborg, D. B. (2004), America's Army: more than a game, Transforming Knowledge into Action through Gaming and Simulation. Ed. & Thomas Eberle Willy Christian Kriz. Munchen:SAGSAGA
- [7] Farrell, C. M. , Klimack, W. K. , and Jacquet, C. R. (2003), Employing Interactive Multimedia Instruction in Military Science Education at the U.S. Military Academy, Interservice/Industry Training, Simulation, and Education Conference
- [8] Reminga, J. and K. M. Carley (2004), ORA:Organization Risk Analyzer, Tech Report, CMU-ISRI-04-106, CASOS, Carnegie Mellon, Pittsburgh PA. <http://www.casos.cs.cmu.edu/projects/ora/index.html>
- [9] Graham, J.M., Schneider, M., Bauer, A., Bessiere, K., & Gonzalez, C. (2004) Shared Mental Models in Military Command and Control Organizations: Effect of Social Network Distance. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society. HFES, California.
- [10] Mike Schneider (2004), Estimating Real-time Shared Situation Awareness using Social Network and Homophily Data, Working Paper, CASOS, Carnegie Mellon, Pittsburgh PA.

- [11] Moon, I. , Schneider M., and Carley, K. M. (2005), Investigation of the Realism of the America's Army Video Game, Working Paper, CASOS, Carnegie Mellon, Pittsburgh PA.
- [12] GlobalSecurity.org, Unit of Action, <http://www.globalsecurity.org/military/agency/army/ua.htm>
- [13] John Graham, Physical Proximity-based Measure of Shared Situation Awareness Mechanisms in Military Network Organizations, NAACSOS Conference, Pittsburgh, PA
- [14] Endsley, M. R., & Jones, W. M. (1997). Situation awareness, information dominance, and information warfare (No. AL/CF-TR-1997-0156). Wright-Patterson AFB, OH: United States Air Force Armstrong Laboratory.
- [15] Albert A. Nofi (2000). Defining and Measuring Shared Situation awareness, CRM D0002895.A1/Final, Center for Naval Analyses, VA
- [16] John Graham (2005). Dynamic Network Analysis of the Network-Centric Organization: Toward an Understanding of Cognition & Performance, Doctoral degree dissertation, CASOS lab, Carnegie Mellon University, PA